基于跨层全连接神经网络的癫痫发作期识别 *

王凤琴 [1]，卢官明 [2†]，柯亨进 [3]，肖新凤 [4]

(1. 湖北师范大学 物理与电子科学学院，湖北 黄石 435102; 2. 南京邮电大学 信息与通信工程学院，南京 210003; 3. 武汉大学 计算机学院，武汉 430072; 4. 广东环境保护工程职业学院，广州 528216)

摘　要：在缺乏足够先验知识下，自适应癫痫发作期识别异常困难。提出一种新的度量通道之间的同步特征计算方法（聚类划分互信息），以相关矩阵方式组织单窗口内全局同步特征模式，进而设计一种跨层全连接神经网络分类器，对非平稳同步特征模式实现自适应分类。实验表明该方法可获得[98.19%±0.24%]精确度，[98.27%±0.51%]敏感度和[98.11%±0.36%]特异度，超过了大部分现有方法的分类性能。另外，所提出方法无须去噪和去伪迹等预处理过程；而且其仅需设置一个超参数（时间窗），避免了过多的潜在错误参数设置而导致的分类性能的降低。

关键词：聚类划分互信息；脑电；癫痫；同步；模式分类；跨层全连接神经网络

中图分类号：TP183　　doi: 10.3969/j.issn.1001-3695.2018.01.0017

Epileptic EEG identification with cross layer fully connected neural network

Wang Fengqin[1], Lu Guanming[2†], Ke Hengjin[3], Xiao Xinfeng[4]

(1. *Hubei Normal University*, *Huangshi Hubei* 435102, *China*; 2. *Nanjing University of Posts & Telecommunications*, *NanJing* 210003, *China*; 3. *Wuhan University*, *Wuhan* 430072, *China*; 4. *Guangdong Polytechnic of Environmental Protection Engineering*, *Guangzhou* 528216, *China*)

Abstract: Under the circumstance of insufficient prior knowledge, it becomes even more important to adaptively classify the synchronization dynamics to accurately characterize the intrinsic nature of seizure activities represented by the EEG. This study first measures the global synchronization by calculating Clustering Partition Mutual Information (MI) of all EEG data channels. A cross layer fully connected net is then designed to adaptively characterize the synchronization dynamics captured correlation matrices and automatically identify the seizure states of the EEG. Experiments are performed over the CHB-MIT scalp EEG dataset to evaluate the proposed approach. Seizure states can be identified with an accuracy, sensitivity and specificity of [98.19%±0.24%], [98.27%±0.51%], and [98.11%±0.36%], respectively; the resulted performance is superior to those of most existing methods over the same dataset. The approach alleviates the need for strictly denoising and artifact removing based on the EEG prior knowledge that is mandatory for existing methods. Only one hyper-parameter need be set manually to avoid getting worse performance because of complex parameter setting.

Key words: clustering partition mutual information; EEG; epilepsy; synchronization; pattern classification; cross layer fully connected net

0　引言

为理解脑功能与脑疾病的内在机制，研究人员往往评估脑电信号之间的同步模式，同时刻画不同脑区之间的交互[1]。到目前为止，大量度量多变量数据间同步强度方法在不同学科中有了长足发展，诸如特征提取[2]、复杂神经震荡网络、神经计算[3]和脑疾病学[4]。

关于脑电信号同步的早期研究专注于双变量同步分析，诸如 Pearson 相关系数、Spearman 秩相关和互信息等。在这些方法中，互信息是一种最重要的信息独立性度量[3]，且在辨别以及抗噪声能力上具有较优的性能[5]。近几年，多变量同步分析方法有了长足发展，诸如相同步聚类分析（PSCA），S 估计子[6]和相关矩阵分析(CMA)[7]。其中，S 估计子能有效度量全局同步，但缺乏对变量间同步细节的度量；PSCA 可以获取不同变量的拓扑细节，但在全局同步信息度量方面存在明显不足；而 CMA 兼顾以上优点[3]。

收稿日期：2018-01-05；**修回日期：**2018-03-03　**基金项目：**国家自然科学基金资助项目（61071167，61501249）

作者简介：王凤琴（1979-），女，湖北黄石人，讲师，硕士，主要研究方向为数字图像处理、机器学习；卢官明（1965-），男（通信作者），江苏南京人，教授，博士，主要研究方向为图像处理、机器学习、计算机视觉、模式识别（lugm@njupt.edu.cn）；柯亨进（1981-），男，湖北武人，博士研究生，主要研究方向为复杂系统分析、深度学习、脑电神经计算；肖新凤（1978-），女，湖南邵阳人，讲师，硕士，主要研究方向为数据挖掘.

序列的划分与互信息的精确计算密切相关。现有的策略主要有两种：a)基于事件的策略，比如计算两个文本之间的互信息，往往基于文本特定含义进行划分来计算自信息熵和联合信息熵，这种方法对度量具体含义的离散随机变量之间的互信息具有较大优势，无法刻画无语义连续随机变量的互信息;b）利用概率分布对连续随机变量进行划分后，再计算互信息。比如在度量两幅图像的互信息时，一般对两幅图像的连续像素值上进行等间距划分（假定服从均匀分布）后再计算自信息熵和联合信息熵，进而计算两幅图像的互信息。上述方法不适合对多变量脑电信号进行分类，其主要原因有：脑电信号本身不具有语义信息，无法通过特定语义进行划分；无明显证据表明多通道脑电信号服从某种特定的概率分布。

癫痫同步模式往往存在非线性、多样性和不确定性，难以通过线性方法识别癫痫发病状态，为此，非线性自适应模式识别技术势在必行。特别是深度学习技术，因为它具有非线性处理能力、自适应性、强大的容错能力。有关神经网络性能的一个误区是神经网络的层数越多，分类器的性能越优。Forrest 提出的 SqueezeNet 具有较少的隐含层和较少的模型参数，却获得了和深度神经网络相同精度层次的分类性能，同时它具有如下优点：可更有效的进行分布式训练，具有较低的负载和容易部署在资源有限的嵌入式平台中。然而，一个层数非常高的神经网络总是面临过分拟合和梯度消失等问题[8]。

针对这些问题，本文首先提出一种能有效抑制强噪声的全局同步特征方法，再设计一种跨层全连接神经网络分类器，对癫痫发作状态进行分类。在 CHB-MIT 公共数据集上的实验结果表明，与大多数已有方法相比，本文所提出的分类器获得较优的分类性能。由于分类器不仅无须足够的先验知识针对特定的病人调整模型参数（去掉病人相关的噪声、特定频率带和眼动等等），而且单一的时间窗口参数设置可大大减少由于参数设置不当而引发的各种错误，致使该分类器可被有效的应用于复杂的科学与工程应用中。其主要贡献如下：

a)针对受到强噪声干扰的多维脑电数据，设计一种提取同步演化模式的特征提取方法（通道间聚类划分互信息），相对于传统互信息计算方法，该计算方法虽然需要额外的聚类过程计算花销（秒级），但其考虑通道数据本身的差异性，进而更加精确的度量双通道的互信息；

b)在缺乏足够的先验知识的情况下，设计一种自适应的区分具有多样性和不确定性同步模式的分类器（跨层神经网络），其在癫痫发作检测中表现出优异性能，相对于现有神经网络方法，跨层神经网络考虑到特征矩阵（互信息矩阵）区别于图像/视频数据元素间的不连续性，去掉可能引起结构信息损失的卷积层和池化层，虽然增加了跨层连接层数而增加时间复杂度，但可避免前层的梯度衰减而缓解梯度消失问题，同时增强跨层之间的特征传播；

c)提出的方法针对整个样本空间，且无须足够的先验知识进行常规脑电信号处理过程中的去噪、去干扰和去眼电过程（这些过程强依赖于病人本身特性），其目的是为了使模型具有更加健壮（无过拟合）和泛化的能力。同时无须对新来的病人重新调整参数，致使模型更具有临床应用意义。

1 相关工作

长期以来，对隐藏在多变量 EEG 中的非平稳模式进行分类受到广大科研工作者的关注，特别是在探索与研究诸如癫痫等脑疾病机理时。传统方法往往侧重于时频分析或者同步度量，然而，随着近年来机器学习方法的蓬勃发展，出现了这一方向的很多显著成果：

在 CHB-MIT 头皮脑电图数据集上，Meyers 等人计算幅度/相位锁定值获得 77%的灵敏度，88%的精确率，以及每小时 0.17 的误报率[9]。为寻找癫痫发作期的起点以及偏移量，Lorena 等人设计开发了一种平稳小波变换的通用癫痫发作期识别算法，获得了 99.9%的特异度、87.5%的敏感度和每小时 0.9 的误报率[10]。Morteza 等人提出了一种基于密度的实时癫痫发作预测算法，获得 86.53%的准确率、97.27%的召回率以及每小时 0.00215 的误报率[11]。

在其他数据集（包括私有数据集）上，Piotr 等人[2]提出一种分类病人相关的同步模式的方法来预测癫痫发作期。在 Freiburg 癫痫数据集上取得了 71%的灵敏度和零误报率。Frèdèric 等人[12]计算左右颞叶与顶叶的脑电癫痫同步，该方法取得了 100%的特异度、54%的灵敏度和 81%的精度。Fergus 等人[13]提出一种新的可以对不同患者进行癫痫检测的通用方法，达到 88%的灵敏度和 88%的特异度。

在国内，基于机器学习的癫痫自动分类也受到广大学者的密切关注。国内在该领域中主要针对德国伯恩大学癫痫研究中心的单通道脑电数据。其代表算法包括小波多尺度分析[14]、可调品质因子小波变换[15]、自回归系数[16]和互信息[17]。在分类器上主要以极限学习机[14]、支持向量机[15,17]和关联向量机[16]为主。在单通道 EEG 数据上大多可以获得 99%以上的分类性能[14-16]，甚至是 100%[17]。在 CHB-MIT 头皮脑电图数据集上,单绍杰等人利用 LSTM 的神经网络模型对癫痫发作进行预测，取得了 98.5%的分类精度和零误警的结果[18]。

脑电信号是一种随机性很强的非平稳信号，具有信噪比非常低的特点，现有方法往往通过去噪和去伪迹等预处理手段，该过程不可避免的需要依赖专家知识，极大地妨碍临床和科学上的应用。与现有的工作相比，本文旨在寻找高噪声鲁棒性的同步模式计算方法，该方法能更加精确的计算互信息，同时设计可增强跨层之间特征传播的跨层全连接神经网络对癫痫发作期自动分类且取得较优的分类性能。

2 方法

2.1 方案概述

考虑到临床应用的时效性和在线性，本文避免采用常规的、繁重的数据预处理任务（去噪和去干扰等）。已有方法影响临床

应用的另外一个障碍是极大的依赖于足够的先验知识，同时需要设置额外的超参数（如噪声和干扰的频率带，空间过滤阈值等）。图 1 显示了本文方法的总体结构图，主要由下面三个阶段：同步特征提取，基于跨层全连接神经网络的模式分类和对所发现的模式进行评估。系统首先对原始数据进行等窗口划分；然后计算每个窗口内所有通道间的聚类划分互信息，并组织成相关特征矩阵；最后利用训练集训练分类器的连接权值并进行癫痫发作期分类。

2.2 一种聚类划分互信息同步方法

1）互信息

互信息是信息论中的基本概念，其不需要任何假设或先验知识就可度量两个随机变量统计相关性。设 X 和 Y 是两个离散的随机变量，其概率密度分布分别是 $p(x)$，$p(y)$，联合概率密度分布为 $p(xy)$，则随机变量 X、Y 的熵 $H(X)$、$H(Y)$ 及其联合熵 $H(X, Y)$ 如下：

$$H(X) = -\sum_{x\in X} p(x)log_2 p(x) \tag{1}$$

$$H(Y) = -\sum_{y\in Y} p(y)log_2 p(y) \tag{2}$$

$$H(X,Y) = -\sum_{x\in X,y\in Y} p(xy)log_2 p(xy) \tag{3}$$

则互信息计算如下：

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{4}$$

2）聚类划分互信息

互信息与数据的划分密切相关，为更加精确的计算互信息，有必要寻求有效地非监督划分策略。聚类算法借助度量特征间的相似性，将物理或抽象对象的集合分成由相似的对象组成的多个类的过程。从集合论的角度看，聚类旨在解决集合的划分问题。比如，基于距离的划分、基于层次的划分、基于密度的划分、基于网格的划分以及基于分布密度的划分等。

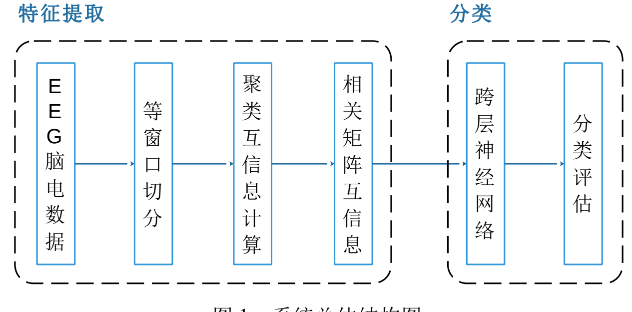


图 1 系统总体结构图

聚类划分互信息可被用来度量双变量脑电数据间的线性和非线性同步关系。相比于现有计算方法，其充分考虑了序列自身的差异性而进行合理的划分，以期精确计算其互信息。图 2 显示了单窗口内聚类互信息相关矩阵计算过程。算法 1 显示了计算两个随机变量之间的聚类划分互信息。

算法 1：计算两个随机变量之间的聚类划分互信息

输入：随机变量 X，随机变量 Y，聚类中心个数 K。

输出：互信息值。

1. 对 X 和 Y 按照从小到大进行排序；
2. 利用 KMeans 算法计算随机变量 X 的 K 个聚类中心以及每个相邻聚类的分割点；假设相邻的两个聚类中心的坐标分别为：$(x_1, 0)$和$(x_2, 0)$，相应的聚类半径分别为 r_1 和 r_2，其划分分割点 S_x 的坐标为 $\left[\frac{(x_2 - r_2) - (x_1 + r_1)}{2}, 0\right]$；
3. 同理，可以计算随机变量 Y 的 K 个聚类中心以及每个相邻聚类的分割点 S_y 坐标为 $\left[0, \frac{(y_2 - r_2) - (y + r_1)}{2}\right]$；
4. 以 S_x 对 X 进行划分 D_X， 对 X 计算每个划分的点的个数 N_X 及其概率 P_X，计算神经信号数据 X 的信息熵 H_X；
5. 同理以 S_y 对进行划分 D_Y， 对 Y 计算每个划分的点的个数 N_Y 及其概率 P_Y 计算神经信号数据 Y 的信息熵 H_Y；
6. 计算 X 和 Y 同时落入划分 D_X 和 D_Y 的个数 $N_{X,Y}$ 及其概率 $P_{X,Y}$；同时计算 X 和 Y 的联合信息熵 $H_{X,Y}$
7. 计算 X 和 Y 的互信息并返回。

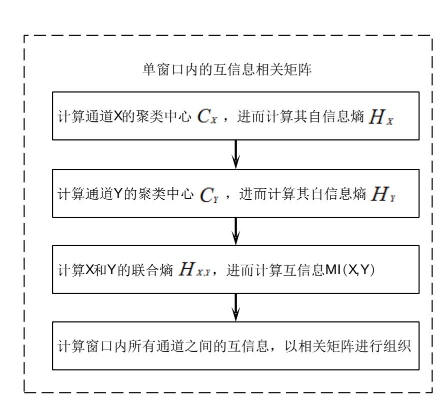


图 2 单窗口内聚类互信息相关矩阵计算过程

2.3 聚类划分互信息相关矩阵

为量化多变量 EEG 的全局同步，本文将所有通道间 MI 组织成一个相关矩阵，即 CMMI。其定义如下：

$$\text{CMMI} = \begin{bmatrix} \text{MI}_{11} & \text{MI}_{12} & \cdots & \text{MI}_{1n} \\ \text{MI}_{21} & \text{MI}_{22} & \cdots & \text{MI}_{2n} \\ \vdots & \vdots & & \vdots \\ \text{MI}_{n1} & \text{MI}_{n2} & \cdots & \text{MI}_{nn} \end{bmatrix} \tag{5}$$

其中，MI_{ij} 表示通道 i 与 j 之间的同步强度。

2.4 跨层全连接神经网络分类器

分类器被用来所有时间窗口内的 CMMIs 进行分类，以对癫痫发作期进行识别。本节先讨论分类器的设计原则，再给出分类器模型的实现细节。

2.4.1 跨层全连接神经网络设计原则

本文分类器的设计目标是用尽量少的隐藏层而获取和深度神经网络同等级别的分类性能，从而节省大量的训练时间。图 3 显示了跨层全连接神经网络的主要结构，其开始于一个 Dropout 层，接下来是四层前向两两相连的密集层块（前面的全连接层都与所有后面的全连接层相连），最后是一个 ReLU 的激

活输出层。其主要设计原则如下：

a）Dropout。Dropout 被用来克服过分拟合的重要手段，其主要思想是随机丢失掉某些神经元。Srivastava 给出了 Dropout 的经验值为输入层为 20%，隐含层为 50% [19]。本文经过大量实验进行参数调优后，设置输入层的 Dropout 率为 20%，而隐含层没有设置 Dropout。

b）无须卷积层和池化层。虽然卷积层和池化层在减少模型参数上取得了极大成功，但其不适合处理 EEG 模式。因为 EEG 模式往往呈现出多样式和非平稳性等特点，卷积和池化操作不可避免的会导致信息丢失而降低分类性能。

c）拼接层。拼接层接受来自所有前层的连接，按照层的前后顺序进行拼接，然后一一映射到接下来的全连接层，其连接权值固定为 1，偏置固定为 0。

d）密集层块。密集层块由块内所有 FC 层之间以前向全连接的方式连接而成。假设有 L 层 FC 层，除了相邻的两层之间都有前向连接之外，所有前面的层都连接到后面所有层，使得神经网络总的连接个数为 $\frac{L(L+1)}{2}$ 层。该网络结构借鉴自 Huang 提出的 DenseNet[20]，文中指出密集层具有缓解梯度消失问题、增强特征传播、更有利于特征重用以及具有更少的模型参数等优点。与其主要不同之处在于：a）由于同步模式的多样性而去掉了卷积层和池化层，因为在不同的癫痫发作期病人，其同步模式增强或者减弱的通道是变化的，臆断的卷积或者池化操作极有可能会损失这种特征的变化而致使最终分类性能降低；b）设置拼接层以接受并拼接所有前层的输出连接，其目的是除了可以避免前层的梯度衰减而缓解梯度消失问题，还可以增强跨层之间的特征传播。

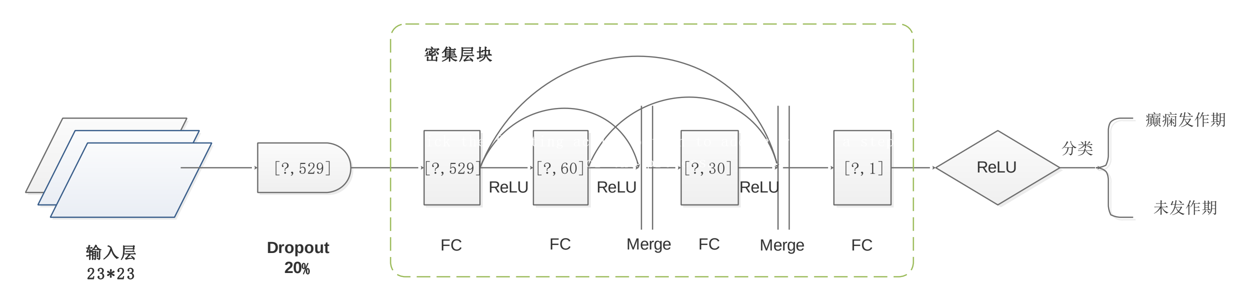


图 3　分类器设计图

2.4.2 跨层全连接神经网络设计结构

分类器详细设计如下：a）优化器,随机批（批次 50）梯度下降算法；b）学习率设置为 0.01；c）目标函数为均方误差；d)激活函数为 ReLU 函数。跨层全连接网络输入所有的 CMMIs，其训练过程包括：

a)前馈网络通过连接拓扑结构和激活函数获得当前层的输出，假设当前层是 L(i)层，除了有来自 L(i-1)层的输入外，还有来自 L(i-1)之前所有全连接层的输入连接，其输出表达式如下：

$$O_{L(i)} = \delta_{L(i)}\left(\sum_{j=1}^{i-1}\delta_{L(j)}\left(\omega_{L(j)}x + b_{L(j)}\right)\right) \tag{6}$$

其中: δ 是激活函数，ω 是连接权值，b 是偏移量。

b）每一次迭代的残差被计算出来；

c）反馈网络依据拓扑结构运用 BP 算法计算输出与目标之间的残差，基于随机梯度下降调节模型连接权重进行优化；由于从输出到当前神经元连接路径的唯一性，反向传播算法依据链式法则一层一层往前计算下降的梯度；

d）训练后得到优化的分类模型，测试数据放入模型后利用已学习参数进行学习特征，最后根据各个特征进行癫痫发作期分类。

2.4.3 分类器实现细节

本节将描述分类器的实现细节。其主要过程包括，首先用随机种子为 7 的随机数对整个样本空间进行重新洗牌，将数据分成训练集、验证集和测试集，其分别占 64%、16%和 20%。分类器的训练性能通过 5 折交叉验证算法进行评估（训练集和验证集），而最终的分类器的性能由测试集进行评估。

a)模型概要。表 1 总结了本文分类器的模型参数（表中的最后一列显示了当前层的参数数目）。虽然分类器的总参数数目（49702）少于当前主流的深度学习框架，但是却获得了较优的分类性能。

表 1　分类器的模型参数设置

层（类型）	输出层	输出格式	模型参数
Dropout1 (Dropout)	FC1（全连接层）	[None 529]	0
FC1（全连接层）	FC2（全连接层）		31800
FC1（全连接层）	FC3（全连接层）		15900
FC1（全连接层）	FC4（全连接层）		530
FC2（全连接层）	FC3（全连接层）	[None 60]	1380
FC2（全连接层）	FC4（全连接层）		61
FC3（全连接层）	FC4（全连接层）	[None 30]	31
FC4（全连接层）	输出（激活函数）	[None 1]	0

表 1 显示了当前层与下一个连接层（输出层）的输出格式以及连接参数个数。密集层块包含两个拼接层，第一个拼接层是在第三 FC 层之前，其拼接来自第一 FC 层和第二 FC 层的输出；第二个拼接层在第四 FC 层之前，拼接来自前面三层 FC 层的输出。

b)训练阶段。在每个训练周期（epoch），5 折交叉验证算法

被用来评估训练性能。采用小批量（批大小为 50）动量（0.9）梯度下降法来对采用的算法进行训练[16]。其正则化的权重衰减值为 0.0001，输入层的 dropout 率为 0.2，设置很小的权重衰减值有助于减少训练误差，其反向更新规则如下[21]：

$$v_{t+1} \leftarrow 0.9 \cdot v_t - 0.0001 \cdot \varepsilon \cdot \omega_t - \varepsilon \frac{\partial L}{\partial \omega} |\omega_{iD_i} ,$$

$$\omega_{t+1} \leftarrow \omega_t + v_{t+1} \quad (7)$$

其中:i 是迭代次数，v 是动量变量，ε 是学习率，$\frac{\partial L}{\partial \omega}|\omega_{iD_i}$ 是目标函数关于连接权值 ω 在 D_i 批上的偏导数，其显示了当前批次的优化方向。

c)测试阶段。在测试阶段，给定一个已经训练好的跨层全连接神经网络和 EEG 原始数据，以如下方式进行分类：首先，EEG 数据被划分为相同大小的数据窗口（本文采用的是 2048），同时每个窗口的 CMMI 特征矩阵被计算出来，最后训练好的神经网络模型对所有 CMMI 特征矩阵进行分类。整个过程无须医生和专家的介入，自动化完成，在临床应用中可以大大节省人力、物力和财力。

d)避免过分拟合。本文使用提前终止和“Dropout”来避免模型的过分拟合。提前终止策略利用迭代截断法在训练收敛到训练集合之前就停止迭代，从而避免过分拟合，本文监控训练精度不再提升时提前终止训练（初始 epoch 设置为 300，而提前终止于 49）。Dropout 是深度神经网络中防止过分拟合提高效果的有力手段。其可以从原始网络中找到更加稀疏的网络，同时减弱神经元节点间的联合适应性，增强泛化能力以避免过分拟合[19]。

3 案例分析—癫痫发病期分类

为评估所提出方法的性能，本文将对麻省理工的公共授权多通道癫痫脑电数据集 CHB-MIT(http://physionet.org/physiobank/database/chbmit[22])进行癫痫发作状态进行分类。

3.1 计算复杂度

本文所有实验的测试环境为英特尔 i7 处理器(3.33 GHz)、24Gb 运行内存和 64 位 Windows7 个人电脑。计算过程包括线下训练过程和线上预测分类过程。

聚类划分互信息的计算复杂度分为两个阶段，其一是 KMeans 聚类算法，对于 K 个划分的聚类算法，其一次聚类划分的计算复杂度为 $\mathrm{O}(KNI)$，N 表示样本个数，I 为迭代次数；其二是互信息的计算，其计算复杂度为：计算互信息时的划分个数和聚类个数相同，为 K，假设双通道之间对应的划分中包含的样本点个数为 N，只需要遍历一次划分内的数据就可以计算相应熵，故一次划分的互信息计算复杂度为 O(N)，因此，互信息的计算复杂度为 $\mathrm{O}(\mathrm{KN})$。一次聚类划分互信息的计算复杂度为 $\mathrm{O}(KNI)+\mathrm{O}(\mathrm{KN})$。假设通道数为 C，总共聚类计算个数为 C，而互信息的计算个数为 C(C+1)/2。综上所述，同步特征矩阵的计算复杂度为 O(CKN(I+C))。

假设神经网络的层数为 L，每一层的神经元数为 U，层与层之间的连接个数为 L(L+1)/2。分类器的时间复杂度为 O(UL(L+1)/2)。

3.1.1 线下训练过程

线下训练过程包括计算所有聚类划分互信息相关矩阵（CMMI）和训练神经网络模型。在时间窗口为 2048 下，聚类操作可以重用，23 个通道的聚类操作只需要 23 次聚类操作，10 个线程同时计算聚类，所有通道总共聚类操作的时间大约为 10 s，276 次互信息计算的时间大约是 40 s。在 JDK1.8 中计算一个 CMMI 所花费的时间大约为 50 s。计算所有的 1406 个 CMMI（753 个癫痫发作，753 个癫痫间歇）所花费的时间大约需要 19.5 h。第二步在 1 分钟内输出模型。

3.1.2 线上秒级实时分类过程

线上预测过程包括计算一个 CMMI 需要 50 s；在 0.01 s 内加载模型文件并预测癫痫的状态。

3.2 数据描述

本实验在 CHB-MIT 头皮脑电数据库上进行评估，数据集采集自 22 名由器质性病变引起的严重癫痫病患者（5 名男性，年龄段 3-22；17 名女性，年龄段 1.5-19），数据通过 23 个不同的通道同时采集（FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FZ-CZ, CZ-PZ, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8,F8-T8, T8-P8, P8-O2, P7-T7, T7-FT9, FT9-FT10, FT10-T8, T8-P8）。通过连接在（受试者）头皮表面的 19 个电极与 1 个接地电极以频率 256Hz 采集数据，大多数数据文件中包括多次癫痫发作。

为避免样本不平衡的问题，利用马尔可夫链蒙特卡罗（MCMC）方法采样以平衡癫痫发作及癫痫间歇期的样本。具体方法如下：a）对于每个癫痫发作期，其发作总时间为 S(seizure)，该发作期的 CMMI 个数为 $\mathrm{count}(\mathrm{seizure})=\mathrm{S}(\mathrm{seizure})/\mathrm{S}(\mathrm{window})$，其中 $\mathrm{S}(\mathrm{window})$ 为时间窗口；b）对癫痫发作期前期进行 MCMC 采样，采样个数为 $\mathrm{count}(\mathrm{previous})=\frac{1}{2}\times\mathrm{S}(\mathrm{seizure})/\mathrm{S}(\mathrm{window})$；c）对癫痫发作后期进行 MCMC 采样，采样个数为 $\mathrm{count}(\mathrm{post})=\mathrm{count}(\mathrm{seizure})-\mathrm{count}(\mathrm{previous})$。

3.3 分类性能评估

为评估分类器的性能，本文拟采用 10 次评估过程，在每一次迭代评估中， 5 折交叉验证方法被采用，所有的输入特征 CMMIs 被随机并打散（shuffle）划分为 5 部分，其中 4 个部分被用作训练数据，剩余的一个部分作为测试数据，最终的结果是所有 10 次迭代所产生的测试集的平均分类结果。为量化分类性能，将利用敏感度（SEN）、特异度（SPE）、准确度（ACC）报告分类器的分类性能。

$$\mathrm{SEN}=TP/(TP+FN) \quad (7)$$

$$\mathrm{SPE}=TN/(TN+FP) \quad (8)$$

$$\mathrm{ACC}=(TP+TN)/(TP+FN+TN+FP) \quad (9)$$

$$\mathrm{Pre} = \mathrm{TP} / (TP + FP) \quad (10)$$

$$\mathrm{GM} = \sqrt{SEN*SPE} \quad (11)$$

$$\mathrm{F_1-Score} = 2 \times (Pre \times SEN) / (Pre + SEN) \quad (12)$$

其中:SEN 和 SPE 分别表示识别癫痫发作期与非发作期的正确率，ACC 表示分类器的平均性能。Pre 定义了所有正确分类中，其正确探测癫痫发作期的百分比。以上指标不能度量分类器的平衡性，例如 100 个样本，如果 99 个正类和 1 个负类，99%的高性能可以被获得即便所有的样本都被分类为正类。Gm 和 F1_Score 综合考虑了不同性能指标而对分类器的性能指标的平衡性进行评估。时间窗口的大小会极大的影响分类器的性能。图 4 显示了不同时间窗口大小下的分类器的性能，随着时间窗口大小的增加（开始于 512），分类器性能的变化趋势呈现上升趋势，一个例外是在时间窗口为 1000 样本点的时候，所有的性能几乎都达到了最低点。而在 2048 样本点的时候，本文的性能达到最高点，也就是本文报告的分类性能：[98.19% ± 0.24%]精确度, [98.27% ± 0.51%]敏感度和[98.11% ± 0.36%]特异度。

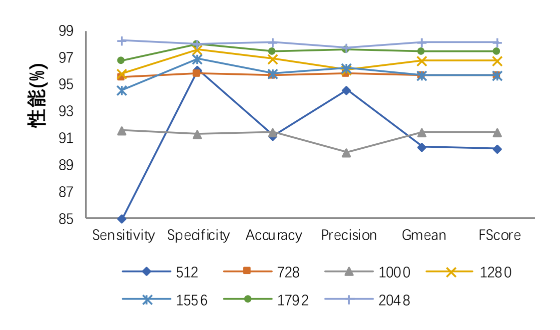


图 4 不同时间窗口大小下的分类器的性能

图 5 显示了 2048 时间窗口下的一次学习曲线，训练精度、训练损失分别表示训练阶段的精度和损失，而验证精度和验证损失分别表示验证阶段的精度和损失。从图中可以看出，模型训练阶段未出现明显的过分拟合，因为（1）训练精度和验证精度最终几乎同时达到较高，（2）没有明显的差异在两个精度曲线之间，虽然验证精度在一定程度上有一定的震荡，但是总的来说并未明显的脱离训练精度曲线。进一步的，本文报告的高性能（低标准差）表明本文分类器未出现过分拟合。

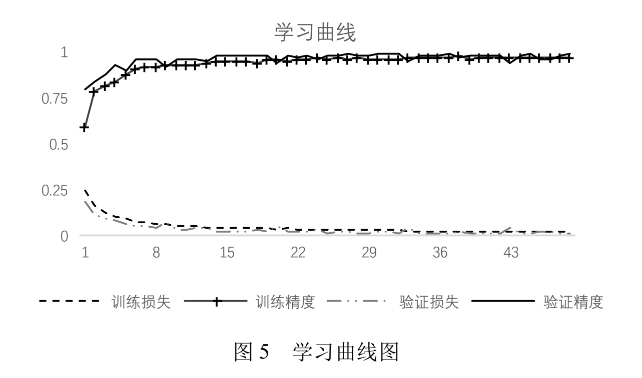


图 5 学习曲线图

受试者工作特征曲线（ROC 曲线）下的面积（简称 AUC）被广泛用来度量分类器的性能。它是反映敏感性和特异性连续变量的综合指标，它通过将连续变量设定出多个不同的临界值，从而计算出一系列敏感性和特异性，再以敏感性为纵坐标、（1-特异性）为横坐标绘制成曲线，曲线下面积越大，诊断准确性越高。在 ROC 曲线上，最靠近坐标图左上方的点为敏感性和特异性均较高的临界值[23]。图 6 显示了本文交叉验证的分类性能（5 折交叉验证）。凸的 ROC 曲线（平均面积为 0.99）显示了分类癫痫发作与非发作状态的优越性能。

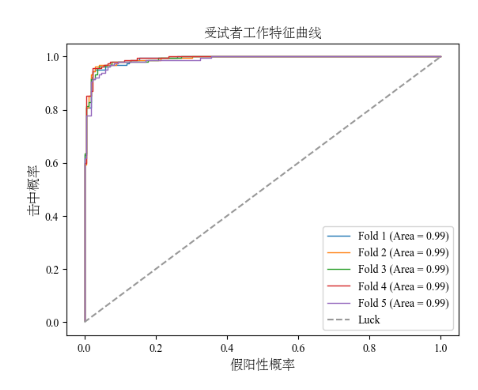


图 6 5 折交叉验证模型的受试者工作特征曲线

Luck 线可以区分模型是否有效，一般来说，如果 ROC 曲线位于 Luck 线以下，可以认为模型是无效模型，反之为有效模型。

表 2 给出了本文方法与当前最具有代表性的智能分类算法的比较。先验知识列表示是否提出的方法强烈依赖于专家的先验知识，诸如利用专家的先验知识过滤与癫痫无关的频率信息等。在公共数据集 CHB-MIT 上，在缺乏先验知识（无须去噪，无须频率带过滤，针对整个样本空间而不是针对单个样本）的情况下，在所有性能指标上都取得了相对较高的性能。这样做的目的是为了使得模型具有更加健壮和泛化的能力，而无须对新来的病人重新调整参数而丧失临床应用意义。

表 2 典型发作期探测方法性能比较

作者/年份	分类器	敏感度/%	特异度/%	精度/%	先验知识
Fergus 2016[13]	k-NN	88	88	93	Y
Nasehi 2013[24]	IPSONN	98	-	-	Y
Morteza 2016[11]	MLP, Bayesian	86.53	97.27	86.56	Y
Lorena 2016[10]	LDA,NN	97.5	99.9	-	Y
单绍杰[18]	LSTM	98.73	-	98.95	Y
本文方法	跨层全连接网络	98.27	98.19	98.11	N

3.4 讨论

大多数已有癫痫发作期分类器都基于病人相关[24]和强依赖于专家的先验知识[10,11,13,18,24]，其缺陷是训练和测试的样本来自于同一个病人，或者针对不同的病人设定特定的特征提取规则，导致较弱的模型泛化能力。作为对照，本研究采用所有患者的样本，在此基础上建立一个通用的 EEG 分类模型，以准确地探

测不同受试者的癫痫发作状态。

传统分类器大多依赖脑电信号的时间、频率和空间分析[2]。而不同病人的频率带常常具有比较大的差异性，而识别一组合适的频率带本身已经成为一个极具挑战的研究方向，致使对不同病人提取不同频率带成分进行分类异常困难。已有方法针对这一问题，例如基于贝叶斯框架提取频率带成分[25]。

另外，提取合适的频率带需要复杂的算法处理和大量的迭代周期，同时时间窗口的长度必须足够长，以避免丢失有用的频率信息的风险。例如，Piotr 等人不得不用 1 到 5 分钟的脑电信号（12~60 帧）来获取合适的频率带信息，但是所需的同步信息提取仅仅需要 5 s 的时间窗口脑电信息[2]。

如上所述，现有的分类器需要足够的先验知识。而本文所提出的方法无须所有先验知识。此外，本文的方法无须现有方法无法跨越的预处理（去噪和去伪迹）阶段。

4 结束语

在强噪声背景下，发现多变量脑电信号的同步模式，并在先验知识不足的情况下准确地对其分类是一个重要问题。这种能力可以极大地辅助癫痫发作期的探测与诊断。

设计一种轻量级的跨层全连接网络自适应地刻画癫痫发作期的非平稳模式，并对其有效分类。区别于以往的分类器，本文设计一种拼接层接受来自所有之前的全连接层的输出，一一映射到下一层，以此构成前馈全连接块。该设计可以减轻梯度消失问题和增强网络连接参数的传输。

在公共数据集 CHB-MIT 头皮脑电数据上的实验结果表明，现有方法在分类性能有一定的改进，获得了[98.19% ± 0.24%]精确度,[98.27% ± 0.51%]敏感度和[98.11% ± 0.36%]特异度。小的标准差表明本文分类器的稳定性。鉴于时间窗口与分类性能的强相关性，有理由相信可以通过深度强化学习技术自适应寻求最优化的时间窗口，使得分类性能进一步提高，其超出了本文的范围，而将作为本文的未来工作。相比之下，本文方法无须去噪而获得比较高的分类性能。此外，该方法仅仅需要一个超参数，有效避免了已有方法由于过度的参数设置而出现的潜在错误。在先验知识不足情况下，本文方法极具对暗含在原始脑电信号的复杂同步模式进行有效分类的潜力。

参考文献：

[1] Gysels E. Phase synchronization for classification of spontaneous EEG signals in brain-computer interfaces [D]. Belgique: Université de Gent, 2005.

[2] Mirowski P, Madhavan D, LeCun Y, *et al.* Classification of patterns of EEG synchronization for seizure prediction [J] , Clinical Neurophysiology, 2009 120: 149-171.

[3] Cui Dong, Pu Weiting, Liu Jing, *et al.* A new EEG synchronization strength analysis method: S-estimator based normalized weighted-permutation mutual information [J] , Neural Networks, 2016, 82 (10): 30-38.

[4] Chen Dan, Li Xiaoli, Cui Dong, *et al.* Global synchronization measurement of multivariate neural signals with massively parallel nonlinear interdependence analysis [J]. IEEE Trans on Neural Systems & Rehabilitation Engineering, 2014, 22 (1): 33-43.

[5] Bonita J D, Ambolode II L C C, Rosenberg B M, *et al.* Time domain measures of inter-channel EEG correlations: a comparison of linear, nonparametric and nonlinear measures [J]. Cognitive Neurodynamics, 2014, 8 (1): 1-15.

[6] Carmeli C, Knyazeva M G, Innocenti G M, *et al.* Assessment of EEG synchronization based on state-space analysis [J]. NeuroImage, 2005, 25 (2): 339-354.

[7] Cui Dong, Liu Xianzeng, WanYou, *et al.* Estimation of genuine and random synchronization in multivariate neural series [J]. Neural Networks, 2010, 23 (6): 698-704.

[8] Iandola F N, Han S, Moskewicz M W, *et al.* Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0. 5mb model size [J] , arXiv: 1602. 07360, 2016: 1-13

[9] Meyers M H, Padmanabha A, Hossain G, *et al.* Seizure prediction and detection via phase and amplitude lock values [J]. Frontiers in Human Neuroscience, 2016, 10 (80): 1–9.

[10] Orosco L, Correa A G, Diez P, *et al.* Patient non-specific algorithm for seizures detection in scalp EEG [J] , Computers in Biology and Medicine, 2016, 71: 128-134.

[11] Behnam M, Pourghassem H. Real-time seizure prediction using RLS fltering and interpolated histogram feature based on hybrid optimization algorithm of bayesian classifier and hunting search [J]. Computer Methods and Programs in Biomedicine, 2016, 136: 115-136.

[12] Zubler F, Steimer A, Kurmann R, *et al.* EEG synchronization measures are early outcome predictors in comatose patients after cardiac arrest [J]. Clinical Neurophysiology, 2017, 128 (4): 635-642.

[13] Fergus P, Hussain A, Hignett D, *et al.* A machine learning system for automated whole-brain seizure detection [J]. Applied Computing and Informatics, 2016, 12: 70-89.

[14] 崔刚强，夏良斌，梁建峰，等. 基于小波多尺度分析和极限学习机的癫痫脑电分类算法 [J]. 生物医学工程学杂志, 2016, 33 (6): 1025-1038.

[15] 贺王鹏，杨琳，王芳，等. 基于 TQWT 的癫痫脑电信号的识别 [J]. 生物医学工程研究, 2017, 36 (4): 346-350.

[16] 韩敏，孙磊磊，洪晓军，等. 基于自回归模型和关联向量机的癫痫脑电信号自动分类 [J]. 中国生物医学工程学报, 2011, 30 (6): 864-870.

[17] 沈洋洋，黄丽亚，郭迪，等. 融合互信息和支持向量机的癫痫自动检测算法 [J]. 计算机技术与发展, 2016, 26 (6): 133-137.

[18] 单绍杰，李汉军，王璐璐，等. 基于 LSTM 模型的单导联脑电癫痫发作预测 [J/OL]. 计算机应用研究, 2018, 35 (11) . [2017-11-10]. http://www.arocmag. com/article/02-2018-11-009. html.

[19] Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: a simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning Research, 2014, 15: 1929-1958.

[20] Huang Gao, Liu, Zhuang, Van der Maaten L, *et al.* Densely connected convolutional networks [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[21] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2012, 60 (2) .

[22] Goldberger A L, Amaral L A N, Glass L, *et al.* Stanley. physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals [J]. Circulation, 2009, 101 (23): 5-18.

[23] Hand D J, Till R J. A simple generalisation of the area under the ROC curve for multiple class classification problems [J] , Machine Learning, 2001, 45 (2): 171-186.

[24] Nasehi S, Pourghassem H. Patient specific epileptic seizure onset detection algorithm based on spectral features and IPSONN classifier [C]// Proc of International Conference on Communication Systems and Network Technologies. 2013.

[25] Suk H II, Lee S W. A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces [J]. IEEE Trans on Pattern Analysis and Machine Intelligence. 2013, 35 (2): 286-299.